U.S. DEPARTMENT OF HEALTH & HUMAN SERVICES

ADMINISTRATION FOR CHILDREN AND FAMILIES

# Secretary of Health and Human Services Advisory Committee

# on

# Head Start Accountability and Educational Performance Measures

## Meeting Proceedings

## June 15–16, 2005

## Table of Contents

## Preface

The U.S. Department of Health and Human Services (HHS), Administration for Children and Families (ACF), Head Start Bureau convened the first meeting of the Advisory Committee on Head Start Accountability and Educational Performance Measures on June 15–16, 2005, in Bethesda, Maryland. The purpose of the Advisory Committee is to help assess the progress in developing and implementing the Head Start National Reporting System (NRS) and provide recommendations for integrating the NRS with other ongoing assessments of the effectiveness of the Head Start program. The Advisory Committee will work in coordination with the existing Technical Work Group (TWG), which helped develop the NRS, and make recommendations for how the NRS can be included in the broader assessment frame found in the Family and Child Experiences Survey (FACES), the national Head Start Impact Study (HSIS), Head Start's Performance Based Outcome System, and the ongoing evaluation of the Early Head Start program.

The Advisory Committee's authority is under 42 U.S.C. 9836A; Section 641(A)(b) of the Head Start Act, as amended. The Advisory Committee is governed by the provisions of Public Law 92-463, as amended (5 U.S.C. Appendix 2), which sets forth standards for the formation and use of advisory committees.

**Members**
Dr. Don Bailey
Dr. Thomas D. Cook
Dr. Victoria R. Fu
Dr. Vera F. Gutierrez-Clellen
Dr. Ron Haskins
Dr. Susan H. Landry (Chair)
Dr. Christopher J. Lonigan
Dr. Donald A. Rock
Dr. Prentice Starkey
Dr. Dorothy Strickland

**Ex-Officio Members**

Wade F. Horn, Ph.D.
Dr. Barbara Broman

At the June 2005 meeting, all Advisory Committee members were present, with the exception of Dr. Thomas D. Cook.

## Executive Summary

The U.S. Department of Health and Human Services (HHS), Administration for Children and Families (ACF), Head Start Bureau (HSB) held the first meeting of the Advisory Committee on Head Start Accountability and Educational Performance Measures on June 15-16, 2005. Dr. Wade Horn, Assistant Secretary, ACF, opened the meeting and explained that the purpose of the Advisory Committee is to assemble experts in assessment to help evaluate the development and implementation of the Head Start National Reporting System (NRS) and recommend strategies for integrating the NRS with other ongoing assessments of the Head Start program. Dr. Susan Landry, Chair of the Advisory Committee, provided an overview of the meeting and explained that the purpose of the agenda topics was to inform the Advisory Committee's work. Dr. Horn discussed the impact of the reauthorization of the Head Start program in 1998 and the increased emphasis on accountability and assessment which led to the development of the NRS. The Advisory Committee was then provided with a variety of presentations relating to the NRS.

- Dr. Nicholas Zill, Westat, provided an overview of the NRS and explained that the NRS provides indicators of the progress children are making on key early literacy and math skills. Dr. Zill's presentation outlined the assessment's purpose, process, components, development, implementation, results, current activities, and future plans.
- Dr. Ruth Hubbell-McKey, Xtria, described the development, features, and data management of the Computer Based Reporting Systems, an Internet-based reporting system used to collect the NRS data.
- Diane Paulsell, MPA, Mathematica Policy Research, provided information on the NRS Quality Assurance Study, which examined how the NRS is administered and implemented, views of local Head Start staff, and avenues for system improvement.
- Dr. Louisa B. Tarullo, Mathematica Policy Research, followed up with an overview of the system improvement component of the quality assurance initiative and explained how the study's findings are used to identify areas of concern, recommend system improvements, and communicate information to the Head Start community.
- Joan Ohl, Commissioner, Administration of Children, Youth, and Families, presented an update on the Head Start program's current activities, including specifics about the NRS, the Head Start Impact Study (HSIS), a new GAO Report on the Head Start program's oversight and financial management, the Program Review Instrument for Systems Monitoring (PRISM), the ongoing efforts to improve Head Start's training and technical assistance, and the potential impact of pending reauthorization legislation.
- Dr. Naomi Goldstein, Office of Planning, Research and Evaluation (OPRE), ACF, reviewed the Head Start research agenda and provided context for subsequent presentations on the types of research that ACF is conducting, including the HSIS, the Head Start Family and Child Experiences Survey (FACES), the Head Start

Quality Research Center (QRC) Consortium, the Early Head Start Research and Evaluation Project, and the Child Outcomes Research Support Consortium (CORS).

- Dr. Nicholas Zill, Westat, discussed the research questions, methodology, and findings of the FACES study. FACES is a longitudinal study that has examined three cohorts to explore the characteristics of programs centers, classes, and the impact of Head Start programs on families and children's development.
- Dr. Ruth Hubbell-McKey, Xtria, reviewed the mission and design of the Head Start QRC and summarized the objectives and interventions studied. The QRC is a study of enhancements to Head Start carried out through partnerships between researchers and local Head Start programs.
- Ronna Cook, Westat, reviewed the research questions, methodology, and preliminary findings of the HSIS. HSIS is a national study designed to investigate Head Start's impact on children's development and the circumstances in which Head Start works best for particular children.
- Dr. Michele Plutro, HSB, concluded the presentations with an overview of PRISM and explained how PRISM is used to monitor grantee performance against program requirements.

The Advisory Committee highlighted a number of issues for follow-up discussions:

- Ensuring that all Head Start-related assessment activities are complementary, minimizing duplication when possible, and determining whether assessment guidelines are necessary for programs.
- Exploring how NRS data may benefit program performance at the center, classroom, and individual level.
- Expanding NRS to include social-development, phonological awareness, and additional parental assessment.
- Considering whether evidence-based curricula should be more widely used, or the feasibility of a national curriculum.
- Exploring interventions to engage and support parents in helping their children learn. (Review existing data on parental involvement and satisfaction in available data sets.)
- Disaggregating data on English language learners from existing studies. Methods of determining acceptable rates of growth for this population, and consideration of logistical challenges of administering assessments to this population.
- Consideration of the utility of NRS data for comparing children with low skill levels in both Spanish and English with children who are English language learners with good Spanish skills.
- Determining how effect size should be measured, exploring issues in defining an acceptable effect size, and discussion of whether achievement goals or benchmarks are also necessary.
- Delineating what characteristics among teachers result in improved outcomes (e.g. education level, performance, etc.)
- Determining who should administer the NRS and whether variability among assessors makes it necessary to standardize the administration of the assessment.

- Exploring what accommodations are currently made for children with disabilities in NRS administration, whether these accommodations affect performance, whether they are appropriate, and whether all children with disabilities are (or should be) assessed.
- Consideration of whether NRS should be extended to Early Head Start.

## Purpose of the Committee

*Wade F. Horn, Ph.D., Assistant Secretary, ACF*

Dr. Horn thanked the Advisory Committee members for their willingness to serve on the Advisory Committee and provided an overview of the purpose of the Advisory Committee on Head Start Accountability and Educational Performance Measures. The purpose of the Advisory Committee is to bring together experts in assessment to provide feedback on the NRS and its progress. The Advisory Committee will also consider how the NRS can be integrated with other Head Start assessment activities such as FACES, HSIS, the Program Information Report (PIR), and locally determined assessments that are conducted every year. ACF wants to ensure that all the assessment activities are coordinated and complementary, and that the findings from these numerous assessment activities can be used to improve the effectiveness of Head Start programs.

The Advisory Committee was chartered in March 2004 for two years. Committee membership was announced in January 2005. Each member will serve a two-year term, unless they choose to resign. In that event, a new member will be appointed to serve for the remainder of the term. HHS can extend the Advisory Committee's charter past March 2006.

The Advisory Committee will make recommendations for Michael O. Leavitt, Secretary of HHS, and Dr. Horn. Dr. Michele Plutro is the designated Federal official, and she will oversee the logistical aspects of the Advisory Committee. Barbara Broman, Acting Deputy Assistant Secretary of the Office of the Assistant Secretary for Planning and Evaluation is an ex-officio member of the Advisory Committee.

The Advisory Committee can meet up to three times a year. Ad hoc committees and subcommittees can also be established. In addition, the Advisory Committee will work with the TWG. The Advisory Committee differs from the TWG in that the TWG is a group of experts convened by the contractor and provides advice on the technical aspects of the NRS. The Advisory Committee's role is to provide recommendations to the HHS Secretary and ACF on a wide range of issues relating to the NRS.

Dr. Horn apologized on behalf of Secretary Leavitt who was unable to attend the meeting due to his attendance at a meeting in Russia.

## Overview of Meeting

*Dr. Susan Landry, Chair, Advisory Committee on Head Start Accountability and Educational Performance Measures*

The purpose of the meeting is to provide Advisory Committee members with information about the NRS and other Head Start-related assessment activities. This information will serve as a foundation for the Advisory Committee's work and help members identify areas where recommendations may be appropriate. The information provided at the meeting will also help members to focus their work and identify areas where more information is necessary.

In addition to the topics already on the meeting agenda, members requested that presentations be made on the Program Review Instrument for Systems Monitoring (PRISM) assessment activities and the local assessment guidelines.

## History of the National Reporting System (NRS)

*Wade F. Horn, Ph.D., Assistant Secretary, ACF*

Reauthorization of the Head Start legislation in 1998 added an increased emphasis on accountability and assessment. The legislation calls for the Secretary of HHS to develop methods and procedures for measuring, annually and over longer periods, the quality and effectiveness of programs operated by Head Start agencies and the impact of services provided through the program to children and families. The results-based performance measures developed should be used by the Secretary to identify strengths and weaknesses in the operation of Head Start programs nationally, regionally, and locally. The 1998 legislation added the assessment of programs at the regional and local level. This assessment at the regional and local level is intended to help identify program areas that may require additional training and technical assistance resources.

The reauthorized legislation identifies four educational performance measures for children participating in Head Start that should be included in the assessment process. Children in Head Start programs should: 1) know that letters of the alphabet are a special category of visual graphics that can be individually named; 2) recognize a word as a unit of print; 3) identify at least 10 letters of the alphabet; and 4) associate sounds with written words.

The reauthorized legislation also calls for additional local results-based performance measures. In addition to other applicable results-based performance measures, Head Start agencies are allowed to establish local results-based educational performance measures.

What these changes suggest is that Congress envisioned that the national Head Start program, at the national, regional, and local level, should assess the progress of children in certain areas. In addition, local agencies should have the capacity to develop locally determined assessment systems to measure the progress of the children in their programs.

In September 2001, ACF began a process that resulted in the NRS. ACF reviewed its various assessment activities and determined that they were not designed to assess the impact of Head Start programs and the developmental progress of children in local Head Start programs. The existing assessment activities did not collect, nor had programs ever been asked to provide, this type of information. Programs had been provided guidance on how to do this type of assessment but it was not known whether the locally developed assessment systems had proven reliability or validity.

From a national program perspective, a common assessment process across programs would help to direct resources in order to help weaker programs achieve better outcomes. A common metric across programs was necessary, since differences in outcomes can be a result of the assessment method and not because children are progressing differentially in programs.

ACF identified a core set of measures that complemented the locally determined performance measures. These four domains and nine indicators were drawn directly from the statute. The core measures focus on indicators of early literacy and numeric skills. They do not focus on social-emotional functioning or health indicators.

For each domain and indicator, ACF identified a readily available measure with established reliability for use with 4 and 5 year-old children. ACF is funding the development of appropriate measures for the areas where measures were unavailable.

To further ensure that the measures were appropriate for the population, a pilot test of the battery was conducted. The pilot test was also intended to assess whether local assessors, such as teachers, could administer the assessments in a reliable and valid way. It also assessed whether the measures were too frustrating for 4 and 5 year-old children in Head Start programs. The pilot test included 1,400 children in 36 diverse Head Start programs. The pilot test indicated that one of the measures was too frustrating for the population so that measure was discarded. The following year, teachers were trained in the administration of the revised battery, and the battery was administered over the course of a year.

One of the issues that arose early in the NRS development process was whether to sample or to use a universal assessment system. Sampling can be used to provide a reliable estimate, provided there is a large enough population to be sampled. Head Start programs range widely in size, from as few as six children to as many as 10,000. If sampling was used in the NRS, the system could not be used by programs to assess at the individual level. ACF decided it was easier and more beneficial to programs to start at the individual level and if necessary, revise the system to use sampling, instead of starting with sampling. This approach seemed to provide the greatest possibility for integration with local assessment processes. Currently, individual scores are not shared with programs, only average scores. At some point in the future, ACF may consider providing individual scores.

## NRS Implementation

*Dr. Nicholas Zill, Westat*

The NRS was designed to address the President's *Good Start, Grow Smart* initiative and the 1998 Congressional mandates for the Head Start program. The NRS provides indicators of the progress children are making on key early literacy and math skills for all local Head Start programs. NRS is designed to supplement, not replace, local child assessment and program self-assessment efforts. This is the first time that consistently collected, comparable measures for all programs are available to Head Start. The intent is that the findings will be used in planning training and technical assistance for the programs. The findings will also be incorporated in future program monitoring.

The skills that are assessed by the test were selected based on the following criteria. The skills:

- Are what Congress and the President expect Head Start children to learn;
- Can be reliably measured in a relatively brief assessment;
- Contribute to achievement in school;
- Can be enhanced by activities in Head Start;
- Are what parents want their children to learn.

*The Assessment Process*

All kindergarten-eligible 4 and 5 year-old children are given the same brief one-on-one assessment. Assessments are conducted in the fall and spring in order to show progress over the course of the year. The assessments are conducted by local assessors. To ensure that the assessors are qualified to conduct the assessments, Westat conducted a three-day, train-the-trainer session for trainers from 2,000 Head Start programs. These trainers trained local staff to conduct the child assessments.

During the assessment, the child's responses are recorded by the assessor on scannable answer sheets. The responses are then reported via an Internet-based, Computer Based Reporting System (CBRS). Identification numbers are developed for all children to maintain confidentiality. Programs may import data from local information-management systems into CBRS so that they do not have to enter all the information.

The local programs report the data to Xtria, the subcontractor that runs the computer-based system and identification numbers are generated based on class rosters. Answer forms are sent to Pearson, another subcontractor, where they are scanned electronically. The information is then sent to Westat, along with the information from the CBRS. At Westat, the assessments are scored and integrated with information from the CBRS. Westat prepares individual program reports for the 1,800 Head Start programs, as well as the Head Start Bureau and regional offices. The reports can be downloaded from the CBRS.

*Components of the Assessment*

The assessment includes: an English language screener that determines if the child knows enough English to continue with the assessment; a picture/word vocabulary assessment; a letter naming assessment involving all the letters of the alphabet in upper and lower case; and a short test on early math skills covering simple addition and subtraction, reading of numbers, relative size judgments, and the ability to make use of graphic and pictorial information involving representation of quantities.

The Spanish assessment is the same. Whether they pass the English screener or not, children from Spanish-language families receive the assessment in Spanish. In Puerto Rico, the assessment is given only in Spanish.

*Development Process*
The TWG guided the development process. Westat used extensive assessment data from as far back as 1997 that had been gathered from longitudinal studies on samples of Head Start children (FACES, Quality Research Center [QRC] Consortium Interventions, and HSIS). Westat had a very large database on how Head Start children had done on various assessments, such as the Peabody Picture Vocabulary Test (PPVT). Using this expertise, Westat selected subsets of items with difficulty levels suitable for Head Start children. Based on experience, Westat knew that there was a great deal of dispersion among Head Start children. To address this, a full range of items had to be included in order to capture the children at the higher levels. A thorough assessment was also conducted of what programs were doing with their local assessments and observational systems. This assessment indicated that many programs were not very proficient in assessing children. Since many local programs lacked assessment skills and the training for assessors was going to be brief, Westat had to develop an assessment that did not exceed the capacity of local programs. Ensuring the reliability of the local assessors was a major concern.

The public had an opportunity to comment on the NRS prior to the first year of implementation. Westat received thousands of comments.

*Field Test*
A national field test of 1,430 students from 36 Head Start programs was conducted in April and May 2003. The field test was designed to test the child assessment and the training-of-trainers approach. Parallel assessments by local Head Start staff and trained and experienced Westat field assessors were conducted. One domain, relating to phonological awareness (Elision task), was removed due to a lack of inter-assessor reliability. There was agreement between the mean scores from the experienced assessors and local staff and similar mean scores from teachers and non-teachers. The language screener worked well with children who did not speak English at home.

*National Implementation*
To prepare for national implementation, a "training of trainers" was held in seven cities and Puerto Rico, with 2,831 local staff members trained and certified as English trainers and 576 staff certified as bilingual trainers. To be certified, trainers had to pass a written test and were observed assessing a child. Bilingual trainers had to take the test and

conduct the assessment in both English and Spanish. Trainers returned to their programs and trained and certified more than 20,000 assessors.

In fall 2003, approximately 429,000 answer forms were successfully processed and 410,000 were processed in spring 2004, making the NRS the largest assessment of preschool children ever conducted. The difference in the number of forms processed was primarily due to children dropping out of the program over the course of the year. Data came from 1,790 Head Start programs. Matching fall/spring assessments were obtained for 343,260 children (80 percent).

Each assessment component is scored separately, and there is no attempt to produce a single summary score for the entire assessment. A goal of the process is to provide scores that are meaningful for the local programs, such as the mean percent of items correct and the percentage of children at different skill levels. With the skill levels, Westat tried to build on the work of the Early Childhood Longitudinal Study to identify proficiency levels meaningful to programs. Item Response Theory (IRT) scale scores were also developed to allow for cross-analysis of programs with similar characteristics.

*Results*

Head Start children showed significant growth in all English language skill areas and the reliability of program-level, fall/spring growth rates based on NRS data was good for both the English and Spanish assessments. Virtually no program showed zero growth or negative growth in English skills. Growth was shown in the following areas: understanding spoken English skills, vocabulary skills, letter recognition skills, and early math skills.

In the Spanish growth analysis, Head Start children assessed in Spanish showed growth in all skill areas and Spanish growth rates were comparable to English rates, except for vocabulary. Growth was greater in programs in Puerto Rico, where the instruction is in Spanish. Some programs on the U.S. mainland showed average Spanish growth rates at or below zero, which means that some of the children in English language programs are losing their Spanish language skills.

*Current Activities*

The second year of the NRS is currently underway, and the numbers are comparable to those from the first year. Improvements have been made in training and assessment procedures in response to feedback from the quality assurance study, the TWG, and local staff. In addition, the distribution of materials has been proceeding in a smoother and more timely fashion. Timely distribution of materials had been an issue in the first year.

Scrutiny of the reliability and validity of the NRS data continues and Westat has completed an analysis comparing FACES to NRS. In addition, a multilevel regression analysis of spring achievement levels and fall/spring growth rates at the program, center, class, and child level has been completed. The model includes 1,800 programs and has provided meaningful results relating to variations in the composition of local programs and operational characteristics such as teacher education level, teacher turnover, and full-

day vs. half-day programs. Independent variables from the CBRS and the PIR were used in the analysis to enrich the information about programs. Westat is also attempting to do a qualitative study of high and low-performing programs on the NRS to determine if this is in accord with perceptions of what are high and low-quality programs.

*Future Plans*
There are plans to expand the NRS to include the domains of phonological awareness, social-emotional development, and child health. For the most part, the child health information, such as height, weight, and immunization status, is already collected by local programs and could be entered into the CBRS. Westat is also looking at alternative assessment techniques, such as using computer-assisted personal assessment, which would allow for more standardization of assessment, quicker reporting of scores, and a broader array of testing procedures. The TWG is also still considering the use of sampling in order to reduce the assessment burden and still obtain reasonably reliable program and center estimates.

*Conclusions*
Despite considerable controversy, the NRS has been implemented and appears to be working reasonably well. While doubts existed about the feasibility of the process, the NRS has been able to conduct in-person assessments of all 4 and 5 year-old children on a limited but important set of early literacy and math skills. FACES indicates that the skills that are being assessed do have predictive value into kindergarten and first grade, and other studies indicate third grade reading ability can be predicted from the types of skills that are assessed by the NRS. The results of the NRS will continue to be scrutinized before they are used for any administrative decisions affecting individual programs. In addition, the Head Start Bureau will be examining the feasibility of expanding the scope of the NRS to encompass other domains and use of alternative methods to improve accuracy and reduce the burden.

Head Start programs have played an important role in the success of the NRS. While many of the programs had reservations about the process, they have worked very hard and have done a very good job in implementing the NRS.

***Questions and Comments***

- *Is information available about the consistency and the context in which the local teachers and staff are conducting the assessment?*
  Some analysis is being done related to this topic. Information is available about who does the assessment, their training and background, and other aspects of the assessment.

- *What are the problems related to chronological awareness and the rater*s?
  The task that was used required the assessor to respond in a complex fashion to provide feedback to the child. A more extensive training period could have addressed this but training for assessors is limited. Westat is looking at ways to address this.

- *Could comparative data across programs be used to identify low-performing programs?*
  Westat is modeling the spring levels of the programs as well as the growth scores from fall to spring. The models using the spring levels are more fruitful in that there is more variance accounted for. There seems to be some measurement unreliability in the fall to spring growth scores, which may be a function of it being first year data. As in FACES, the biggest differentiation in the children's scores has to do with the characteristics of the population being served. Programs that are serving children in greater need have spring levels that are lower than the programs that are serving a comparatively more advantaged population. Children that have a second year of Head Start come out significantly higher in the spring, even though the growth rate for those second-year children shows a deceleration. Children in full-day classes make greater gains than those in half-day classes. Also, having a teacher that has a B.A. or A.A. degree is associated with a small but meaningful increase in scores. Less teacher turnover also resulted in greater student advances. The data seem to indicate that there is a system-wide need for the adoption of methods that have proven efficacy in moving these indicators in significant ways.

- *Can the NRS be used to identify low-performing programs?*
  The answer is not always easy, and there are not always resources to draw upon that can enhance programs. Some of the programs that have the lowest performance are in areas of significant need, such as rural areas, where there is a limited pool of resources.

- *Will an observational component be added to the existing process?*
  A future scenario includes coordination between NRS and the system that is now in place. The periodic monitoring that now occurs could also be coordinated with NRS data.

- *What level of data is available to programs?*
  Even though individual, class, and center data are available, the only reports that are provided to programs are at the program level. There is a lot of interest in center-level data, especially for large centers. Providing these reports is feasible. Many programs want class and individual-level data. Some programs made copies of their scoring sheets and scored the assessments themselves to see how they compared with the local assessments, but this was not an authorized activity.

- *Is the PPVT assessment the most appropriate way to measure vocabulary since the instrument is not very sensitive to interventions?*
  Programs should not be expected to bring about a great change in vocabulary since evidence indicates that much of the change is dependent on what goes on at home. However, the PPVT score is a measure of readiness. While vocabulary is hard to change, programs must work toward this. A useful tool might be a pool of

words (1,000 or more) that all children in Head Start should know before they enter kindergarten, and teachers could teach toward this broader set of words.

## Computer Cased Reporting System (CBRS)

*Dr. Ruth Hubbell-McKey, Xtria, LLC*

The CBRS is designed to collect the descriptive information relating to children, classes, teachers, assessors, centers, and programs for the NRS. It is a Web-based system that is accessed by each participating program. Each program can only access their data. The system assigns unique identification numbers to all children and assessors to maintain confidentiality. CBRS allows for the tracking of information across each data collection period so that there can be intervention with programs that are not making progress. The system allows for the distribution of reports to the programs as soon as the data collection period is over. Reference tables are also available.

*Development of the CBRS*

The basic requirements for the CBRS were determined by the Head Start Bureau. Focus groups were held in 2003 to provide input into the capability of programs to use Web-based systems and whether programs had Internet access. The data elements were developed with input from the focus groups, the Head Start Bureau, and the TWG. A Web-based version was developed for the field test. The original plan included development of a software version for the full rollout, but it was not necessary since the Web-based version was so successful. Head Start staff received both classroom and hands-on training in use of the Web-based version.

*CBRS Features*

Important features of the CBRS include: confidentiality and security with a password system and encryption of children's names; an intuitive and user-friendly system, availability of a paper version of the system for programs without computers or Internet access, technical assistance via a helpline or e-mail, and data import and copy over features to reduce the burden on programs. Over 1,300 programs use the copy over features. Ninety programs have used the data import feature, which is especially beneficial to large programs.

*Data Management*

There were over 28,000 users of the CBRS representing 1,800 Head Start programs in 2003–2004. Information about 39,000 teachers, 29,000 assessors, 41,000 classes, and 410,000 children was collected in spring 2004. The children's dates of birth were added to the scoring sheet to enhance the matching of the data, since sometimes the programs were recording the identification numbers incorrectly.

*Additional Benefits of the CBRS*

Additional benefits that had not been anticipated resulted from the implementation of the CBRS. For example, a master list of Head Start grantee numbers was developed. It also serves as a master list for center addresses, which aids in the shipping of materials. It also helps to coordinate the Head Start Bureau's multiple data sets.

*Data from Year One*

There is great range in the size of the programs. Programs range in size from one to 145 centers with a mean of 9.7 centers and a median of six centers. The majority of programs (37 percent) have between 151 and 500 children and another 25 percent of programs have between 50 and 150 children. Programs with fewer than 50 children make up 18.7 percent and programs with more than 1,000 children make up 2.5 percent. The vast majority of providers are centers (87.2 percent), with 7.2 percent as child care centers, 2.7 percent as home visitor clusters, and 1.9 percent providing family child care.

For assessment in English, the majority of the assessors were Head Start professionals (42.9 percent), followed by teachers (35.1 percent), other Head Start staff (17.1 percent), consultants (4.5 percent), and graduate students (.3 percent). For the Spanish assessment, Head Start teachers made up the majority of assessors (44.2 percent), followed by other Head Start staff (27.3 percent), Head Start professionals (24.9 percent), consultants (3.5 percent), and graduate students (.1 percent). There had been concern at the beginning of the NRS process that the assessment would be very burdensome for teachers and programs to make efforts to find outside assessors. Spanish language programs, however, had a harder time finding people who were qualified to do the assessments. These programs were more likely to use teachers.

The average age of children participating in the NRS was 5.1 years at the time of the spring assessment. English-language learners made up 25 percent of the children, and 72 percent of the children represented racial/ethnic minorities. Nine percent of the children had disabilities, which increased over the year as children were identified. Forty-nine percent of the children were assessed in their first year of Head Start. Forty-eight percent of children attend a full-day program. The vast majority of students speak English (74.5 percent), 22.1 percent speak Spanish, and 5.3 percent speak another language (e.g., Far Eastern, Pacific Island, Middle Eastern/Indic, Western European/Slavic, etc.).

*Summary*
The CBRS system has functioned well and was successfully used by Head Start staff. The system has provided additional data and data management and coordination benefits beyond original expectations. The descriptive data are used for the contextual assessments across programs.

***Questions and Comments***

- *Do the data include a detailed description of curriculum?*
  The CBRS is not collecting this information, but it is in the PIR.

- *Do teachers enter the data for the children that they test?*
  All the assessments are scored and then all the scoring sheets are scored and entered in the database. The descriptive elements are entered by a variety of staff, primarily IT staff. The descriptive data are not necessarily entered by the assessor.

- *Are data available on the number of days that a child is in the program?*

The child's entry date is recorded so that it is possible to determine the length of time they have been in the program.

## NRS Quality Assurance Study

*Diane Paulsell, M.P.A., Mathematica Policy Research, Inc.*

The Head Start Bureau contracted with Mathematica to carry out two tasks in support of the NRS effort. The first task is the quality assurance study, which is an independent, third-party assessment of how well the NRS is being implemented by programs around the country. The second task involves working with the Bureau, the TWG, and the other contractors to assist in the process of system improvement and identify possible changes and enhancements.

The quality assurance study examined how well a sample of assessors administered the NRS assessment and how programs have implemented the NRS. The study also explored the views and concerns of local Head Start staff on the usefulness of the NRS data. Part of the process involved looking at the implications of the quality assurance study's findings on system improvement. An additional component of the quality assurance effort involved exploring the implementation of the NRS in migrant and seasonal Head Start programs.

*Data Sources and Methods*

Mathematica visited a nationally representative sample of 35 Head Start programs in the fall 2003, spring 2004, and fall 2004. Within the programs, a random sample of 10 children was selected, and their assessments were observed. A total of 350 assessments were observed. In addition, interviews were conducted with Head Start directors, lead NRS trainers, and lead NRS data managers, and focus groups of assessors were held at each site. Local assessor trainings were also observed. Mathematica has just completed a fourth round of site visits for the spring assessment, and the findings should be available soon.

The criteria and the scoring procedures for the certification process were used as a guide to determine the quality of the assessments. A certification score was computed for each of the assessments observed. A score of 85 is the minimum score to achieve certification. The study found that the mean English certification scores for the fall 2003, spring 2004, and fall 2004 were all over 90. The percentage of English assessments with a score of 85 or higher were 73 percent in fall 2003, 87 percent in spring 2004, and 79 percent in fall 2004. These findings indicate that Head Start staff can effectively administer the assessments in a standardized way and according to the specified procedures. The error rates were similar in spring and fall 2004 and lower than in fall 2003. Errors that were assessed included: straying from the script, coaching; non-neutral encouragement, incorrect hand gestures, pronunciation/inserting articles, and scoring errors. The overall pattern for the errors showed a decrease between fall 2003 and spring 2004, especially in straying from the script and scoring. The pattern held for fall 2004, except for an increase in coaching errors. Scoring errors are fairly low, and the assessors scored 97 percent of the items correctly.

During each round of visits, approximately 45 Spanish assessments were observed. Overall, the quality was very similar to the English assessment, and the errors generally followed the same pattern. The mean Spanish certification scores were 84 in fall 2003, 94 in spring 2004, and 95 in fall 2004.

*Assessing Special Population*

About 70 percent of the programs assessed children with limited English skills. Across the sites, children spoke a wide variety of languages including Vietnamese, Chinese, Arabic, Haitian, and Hmong. There were some difficulties in determining home language, and there were sometimes conflicts in parental reports about the home language. Almost all of the children passed the English screener, but then many had trouble with the assessment. This may indicate that the screener is too easy.

Ninety percent of programs assessed at least one child with a disability, and accommodations were made for children with disabilities. Accommodations include: more time to complete the assessment, being able to complete the assessment in more than one sitting, asking children to repeat answers, redirection and refocusing of children, and allowing children to stand or move around. Programs reported that they were able to assess children with disabilities but some staff reported that they were unsure about which children should be assessed and what accommodations could be made. Staff requested more guidance on assessing children with disabilities and on which children could be exempted from the assessment.

*Areas for Improvement*

The study identified aspects of the process where staff may need more guidance and training. For example, about a third of the programs had trouble finding an appropriate area in which to conduct assessments since they were using all available center space for classrooms. Other areas include: coaching, especially during the Simon Says section; offering neutral encouragement when children are tiring or getting off task; using correct hand gestures for each item; and correctly scoring the counting items in the early math section.

*How Assessors Were Trained*

Initially, staff attended training-of-trainer conferences, where they were certified on the assessment. These trainers then provided training to staff in their programs. In fall 2004, about half of the programs visited needed to train at least one new assessor. All the programs were instructed to provide refresher training for experienced assessors. This consisted of reading the training script, viewing and discussing a video, and role playing. For the English assessment, about 40 percent of the programs followed the suggested refresher format. About 20 percent of the programs followed the suggested format for the Spanish assessment. Most of these programs had the Spanish assessors attend the English refresher and watch the Spanish video. This may have been due to the small number of Spanish assessors at programs and the limited availability of Spanish-speaking trainers and/or trainers certified in the Spanish version.

*Local Program Concerns about the NRS*

The concerns about the NRS expressed by Head Start staff have been consistent across all the site visits. Concerns include: how the results will be used, especially if they are used to make funding decisions; whether the results accurately reflect both the children's and the program's performance; staff time and resources required to implement the NRS; and the possibility that programs might begin to "teach to the test." Concerns about the Spanish version relate to word usage (dialect and regional terms), differences in items across the English and Spanish versions, and the inclusion of additional letters in the Spanish version.

Head Start program staff found the baseline report clear and easy to understand but they reported that their local assessments are more useful. Information from the local assessments is available in a timely manner and the results can be broken down by center, classroom, and individual child. Nearly 80 percent of the programs said that the NRS reports were received too late to be of use in planning services for the children that had been assessed. However, 60 percent reported that they had some plans to use the results. Programs hoped to use the results to identify weak program areas, compare NRS data to local assessments, inform teacher training, and assess curricula. About 30 percent of programs said that the NRS should be combined with the local assessment to ease the burden on programs and about 30 percent of programs recommended the use of alternative methods and materials in the NRS assessment.

Programs reported making changes to their programs in response to the NRS. Thirty-five percent of programs reported that they had made some changes in classroom practices, such as increased focus on letter naming, vocabulary, and graphing and measurement. Twenty-four percent reported that they had purchased new classroom materials.

*Local Program Suggestions for Improving the NRS*

Programs suggested the following changes to the NRS process: improved communication on the purpose of the NRS; addition of at least one domain (usually social-emotional awareness) while not lengthening the overall assessment; more training on specific topics, especially behavior management during the assessment, neutral encouragement, and assessing children with disabilities; combining the NRS with the local assessments; and use of alternative methods and materials.

*Important Implication of the Quality Assurance Study*

The findings of the study indicate various areas where the NRS can be enhanced. Increased communications about plans for the NRS could address the uncertainty on the part of local programs. This information should address how the results will be used and how the NRS was developed. This information should be available in written materials that can be shared with parents.

Reports on the NRS results should be provided to programs in a timely manner so that they can be used for the next program year and reports would be more useful if they were provided at the center level. Programs also need more guidance on how to use the reports in combination with their local assessments. Additionally, programs would like more

resources and support for making appropriate program improvements to address areas where children did not perform well on the NRS.

Based on field experience, Mathematica suggests adjusting the training agenda to make the refresher training shorter and more focused on changes in the assessment. The assessors guide should be referred to more often during the training process, and assessors should be encouraged to read the assessors guide. Assessors should be provided additional materials and guidance in some areas such as providing neutral encouragement and managing the behavior of children during the assessment. These changes to the NRS training could help staff feel more confident and conduct the assessment in a smoother and more accurate manner.

Programs also need more support in assessing special populations, especially English language learners and children with disabilities. They need more guidance in determining the home language, and the assessment should be more flexible to address Spanish dialects and regional terms. Guidance is necessary on the Bureau's expectations for growth in Spanish skills when Spanish is not the language of instruction. Programs also need more guidance on assessing children with disabilities.

Possible improvements to the assessment battery have also been identified. The assessment battery could be improved by: adding a social-emotional development measure; using manipulatives for some of the math items, especially the counting item, to make them easier to score; eliminating the pie question in the math section; and using alternative approaches to the letter naming task, such as allowing assessors to ask the children questions about each letter.

***Questions and Comments***

- *How is home language identified?*
  All children who speak Spanish are assessed in Spanish and English. In the fall, since the children are new to the program, programs rely on the parent reports at intake of home language. The first year, the assessment was first administered in English. Programs suggested that in the second year the test first be administered in Spanish, since the children often had a sense of failure when they did not do well on the English assessment. This can be problematic if the child does not speak Spanish very well. Programs need more guidance relating to this process.

- *Do teachers or people from outside the program do the assessments?*
  Assessors, since they are working with young children on a one-on-one basis, need a unique set of skills, and programs struggled with who should administer the assessments. Some programs thought teachers would be most appropriate since they are better at managing the children's behavior, and the children are comfortable with them. Other programs did not want teachers administering the test because they feared that the teachers would begin "teaching to the test" or that it would be too much of a burden for the teachers. Some programs brought in

outsiders as assessors. There was no difference in the certification scores between teachers and non-teachers.

## Support for System Improvement

*Louisa B. Tarullo, Ed.D., Mathematica Policy Research, Inc.*

The goals of the quality assurance initiative are to:

- Provide the Head Start Bureau with an independent, objective evaluation of the quality of the NRS implementation.
- Recommend ways for further developing and improving the system and its various components.
- Communicate findings to the Head Start community.

The system improvement component of the quality assurance initiative is a very collaborative process. Mathematica is working with the TWG, other contractors, Head Start programs, and the Head Start Bureau.

The TWG has provided ongoing guidance throughout the development and implementation of the NRS. The TWG was formed in 2002 and has met seven times to make recommendations on the selection of the assessment battery, the process of administering the assessments, and the process for communicating the results. It is made up of 16 leading experts in measurement, psychometrics, content areas, and language and cultural issues.

Communicating information about the NRS to the Head Start community and policymakers is a major focus of the quality assurance initiative. One of the information sharing strategies that has been employed is a series of interactive satellite broadcasts addressing various topics relating to the NRS. This format allows for questions from viewers. Other approaches are also used to share information.

The information collected from the quality assurance study is used by the TWG as it considers system improvements, identifies areas of concern, and determine areas where more information may be necessary. The findings have been used to make changes on a fairly rapid basis. For example, based on the findings of the quality assurance study, training approaches and the way certain items are configured have been changed between the fall and spring assessments in a single year.

The TWG has made specific recommendations to the Bureau about the NRS process.

- In October 2004, a subgroup of the TWG, additional experts, and Head Start program staff met to explore the possibility of assessing children in languages other than English and Spanish. The group recommended that the assessments should be provided in the language of instruction, except for Spanish speaking children. Assessments are done in both Spanish and English for Spanish speaking children.

- The TWG is examining relationships between NRS data and research data from FACES, which is a nationally representative sample over several cohorts. The FACES study covers the same period as NRS, so parallel program data at the individual level can be examined and possible predictions to kindergarten can be made. The comparison with FACES allows for greater examination of the validity of the NRS. FACES is administered by trained assessors.

- The Bureau is considering conducting a field test to expand the battery to include social-emotional development. The field test would use teacher ratings of children's social-emotional and learning behaviors. There are several measures that have been tested extensively, either through FACES or HSIS, which could be used in the field test. The field test could also explore modification to the cognitive battery and the use of program-supplied data about children's health and well being.

- Sampling strategies are being considered in order to reduce the burden on some of the larger programs while still providing data at the center level to assist in the program planning process.

- Guidance for programs in interpreting reports relating to NRS and the local assessments should be developed so that the data is useful to local programs.

## Head Start Updates

*Joan Ohl, Commissioner, Administration on Children, Youth and Families (ACYF)*

Commissioner Ohl provided an update on the Head Start program. Head Start is now in its 40th year and Early Head Start is in its 10th year. The Bureau will be celebrating both these milestones.

The implementation of the NRS is one of the most significant of the Bureau's current activities. NRS is an important new tool in that it provides comparable data about the progress of the children in programs across the country. This information will supplement local assessments and will drive training and technical assistance.

Another important assessment activity is the HSIS. This first report on the study was released June 9, 2005. The HSIS is a congressionally mandated, longitudinal study of the impact of Head Start on approximately 5,000 newly entering, 3 and 4 year-old children. The survey is conducted across 75 grantees. Data collection began in fall 2002 and will continue through spring 2006.

The General Accountability Office (GAO) has released a report on oversight and financial management of the Head Start program. The report recommends that ACF develop a comprehensive risk assessment of Head Start programs and improve the process used for collecting information about program risk. Specifically, GAO recommended that ACF be able to identify and address weaknesses that limit programs' abilities to achieve their objectives. The report also recommended that ACF make greater use of its authority to re-compete grants that are awarded to poorly performing grantees. ACF does not believe that it currently has the authority to do this but this may be addressed in reauthorization.

As part of overall monitoring efforts, ACF has established minimum qualifications for all PRISM reviewers in the areas they are reviewing. Reviewers must be certified on an annual basis. Additional reviewers are being recruited, trained, and mentored. A formal assessment process has been implemented in which the Federal team leaders and the reviewers assess the performance of team members after every review (reviews of each grantee are done every three years). Last year, intensive, multi-day trainings for reviewers were conducted. These addressed fiscal and program management and early childhood development. These changes have resulted in a pool of qualified, skilled reviewers that can assess both management and the quality of the Head Start programs.

In another quality assurance initiative, trained reviewers lead teams in conducting re-reviews of a sample of recently monitored programs. GAO was concerned about consistency across reviews. With the re-review, the reviewers go onsite to programs that have been monitored in the past few months and a second, complete monitoring takes place. The results are then compared to the results from the first review to check reliability of current review teams and to see if the findings are consistent.

An additional quality assurance piece is designed to enhance program viability. The fiscal checklist is used by all fiscal reviewers. Grantees that appear to be facing current or future problems will undergo a more rigorous review of their fiscal systems and records to determine if there are operational or management problems.

ACF is also focusing on training and technical assistance. The training and technical assistance system was reorganized over a year ago. To improve the quality of training and technical assistance, ACF hired contractors specializing in training and technical assistance that operate out of the regional offices. The specialists work on a regular basis with individual grantees. The specialists help grantees determine the training and technical assistance needs and identify appropriate strategies for addressing them. Each grantee has their own training and technical assistance plan, which considers any corrective actions that have been identified in the review process. This process includes multiple site visits. In addition to the local specialists, there is a team of content specialists in each of the regional offices.

Reauthorization is an important issue for ACF. The Head Start Act was due to be reauthorized in September 2003. Bills were introduced in both the House and the Senate this spring, which now must work their way through the legislative process. One of the most significant proposed changes relates to establishing a five-year grant cycle for Head Start grants. Currently, grantees are assured of keeping their grants unless they are found to be deficient. Proposals in both the House and the Senate make it more efficient to replace a poorly performing grantee since grants would be re-competed.

Both the House and the Senate also address teacher qualifications. Each require that 50 percent of teachers in center-based programs have at least a B.A. degree in early childhood education or in a related field combined with experience teaching preschool. Of the 5,600 teachers in the 2003–2004 program year, 30.5 percent had A.A. degrees, 30 percent had B.A. degrees, and 4.3 percent had master's degrees. Prior reauthorization required that at least that 50 percent of teachers had an A.A. degree or higher.

There are provisions in both the House and the Senate bills to raise the funding set aside for Early Head Start to above the current rate of 10 percent. The Senate set-aside goes as high as 18 percent by 2011.

Both bills would require a certain percentage of any increase in funding to be spent on quality improvements. These could include training, course work, compensation, renovation, safety, and other options.

Another proposed change relates to the standards of participation. The Senate bill raises the income threshold from below the Federal Poverty Line (FPL) to below 130 percent of the FPL. Both of the bills expand the list of items not to be considered as income when considering eligibility to include housing benefits or special pay for members of the military.

Both bills require a study by the National Academy of Sciences to make recommendations on appropriate academic requirements, outcomes, and services to ensure that children are school-ready.

Neither of the bills contains a provision for a demonstration program for the integration of Head Start into State education systems. The Administration has proposed a pilot program in which up to nine states would conduct such studies.

Finally, the House bill calls for each state to develop an early learning council to advance and coordinate early childhood learning. The Senate bill calls for up to 200 Head Start agencies to be designated as centers of excellence. Bonus grants would be awarded to these centers, and they could provide services to additional children, model and disseminate best practices, and provide training and technical assistance to improve quality.

## Overview of the Head Start Research Agenda

*Dr. Naomi Goldstein, Office of Planning, Research and Evaluation (OPRE), ACF*

Dr. Goldstein provided context for the subsequent presentations by providing information on the types of research that ACF is conducting.

The HSIS is a nationally representative experimental design, making it a very unique study. It addresses two main questions: 1) what is the overall impact of Head Start on children's development; and 2) under what circumstances does Head Start work best and for which children. The study begins with 3 and 4 year-old children and follows them through first grade.

FACES is a descriptive, longitudinal monitoring study. ACF is now conducting the third cohort of the study. It is a large sample that looks at the characteristics of programs, centers, and classes, and the impact of Head Start programs on families and children's development from Head Start through kindergarten.

The Head Start QRC Consortium is a study of enhancements to Head Start carried out through partnerships between researchers and local Head Start programs. The model has been used to develop, test, document, and evaluate enhancements to Head Start services. The QRC Consortium includes the collection of uniform data across the sites.

Similar assessment efforts are also being conducted in Early Head Start. The Early Head Start Research and Evaluation Project explores the overall effectiveness of Early Head Start and is now following the children in their pre-kindergarten year. This study was not nationally representative.

There are several other university partnership efforts that may be of interest to Advisory Committee members. The Child Outcomes Research Support (CORS) Consortium is developing model approaches for local programs to incorporate assessments of child outcomes into program management and instruction. There is also a set of university partnerships to develop new measures for assessing children's outcomes. Finally, efforts are focusing on special populations, such as migrant and tribal programs. These programs offer special challenges, both for monitoring and research.

In closing, Dr. Goldstein emphasized that both the HSIS and the enhancement projects using experimental evaluations allow ACF to explore the impact of Head Start on children's development. FACES and the NRS have very rich data that show how children are progressing in the Head Start program and how they are doing in comparison to the national average. They do not, however, show the causal impact of Head Start. Cross analysis will allow for the more effective evaluation of Head Start programs.

## Overview of the Head Start Family and Child Experiences Survey (FACES)

*Dr. Nicholas Zill, Westat*

FACES is a mechanism for generating longitudinal findings and secular trend data on the program performance of Head Start. The study results from the Government Performance and Results Act. The effort began in 1996 and 1997 and was based on a series of longitudinal cohorts. Each of the cohorts is based on a stratified national probability sample of Head Start programs, families, and children. Native American, migrant, and Puerto Rican programs are not currently included in FACES. The assessment includes direct assessments of children at the start and end of Head Start, and the end of kindergarten.

As part of the FACES study, three cohorts have been studied to date. The first, beginning in the spring of 1997 and going through 2001, included 40 programs, including Puerto Rico. This cohort was followed through first grade. The 2000 cohort was a sample of 43 programs, and children were followed through the end of kindergarten. The most recent cohort will be completed in fall 2003.

FACES addresses the following research questions.

- What school readiness skills do children have when they enter the program?
- What gains do children make in Head Start?
- How do gains compare with developmental norms?
- What changes in social skills and problem behavior occur while children are in Head Start?
- What is the typical quality of Head Start classrooms as early learning environments?
- What specific curricular approaches are being attempted in Head Start?
- What is the relationship of general classroom quality to children's development gains?
- What is the relationship of the curricular approach to children's gains?

Various measures are used to assess the programs. For example, ratings from parents, teachers, and assessors are used to assess social skills and problem behavior. In the first cohort, classroom observations were also used. Well-established measures of program quality, such as the Early Childhood Environment Rating Scale, are used to assess the classroom environment. The measures that have been used show high reliability and validity and are predictive of children's reading skills and general knowledge at the end of kindergarten.

Multiple methods are also used to assess child development. Theses include:

- Direct one-on-one assessment of children;
- Observation of teacher/child interaction in the classroom;

- Collection of developmental reports from teachers and parents;
- Collection of behavior ratings from teachers, parents, and assessors.

To learn about the programs, the researchers use the following methods:

- Interview center directors, education coordinators, and teachers;
- Examine facilities and contents of classrooms;
- Observe the classrooms in operation;
- Interview parents at the beginning and end of the year concerning their activities with their children and their satisfaction with Head Start.

Demographic and family information is also collected during the parent interviews.

As part of FACES, more than 8,500 children, ages 3 to 5, from 143 Head Start programs have been assessed by trained assessors. Programs have been extremely cooperative throughout the process and the response rates from parents, children, and teachers have been high. Response rates drop off some during the kindergarten follow-up. This is primarily due to the transitory nature of low-income families, not a lack of cooperation.

The analytic approach uses multifactor, multilevel models to depict the relationships between program, class, and child and family characteristics, and school readiness outcome measures. Program quality has also been modeled to determine how it relates to program characteristics and teacher qualifications. Models of the gain in readiness between fall and spring are also used.

*FACES Findings*

The findings on program quality indicate that, as an early childhood learning environment, the average Head Start classroom is of good quality. The range of variation in quality is relatively narrow, especially when compared to some other studies. Very few classes are judged to be of "inadequate" quality. There are also very few that are judged to be excellent.

The findings on children's skills indicate that children enter Head Start with skills that are substantially below the national norms, which is consistent with previous studies. For example, vocabulary scores are at least a standard deviation below the mean. However, the vocabulary scores vary widely—the top 25 percent of children are near the national norm while the lowest 25 percent are close to two standard deviations below the mean. The children make modest gains toward the norms during the program year but remain below the norms at the end of the year.

Additional findings include:

- Fall–spring gains in letter/word identification have increased across the three cohorts of FACES, but no similar increase was found for vocabulary gains. This seems to indicate that the program's focus on literacy skills and possibly even the introduction of the NRS have resulted in some gains in letter knowledge.

- By the end of kindergarten, the average Head Start graduate comes close to national norms in "decoding" or "inside-out" skills like recognizing letters and letter sounds and writing letters on demand. However, the average Head Start graduate continues to be significantly below the national norms in "outside-in" skill areas, such as vocabulary, general knowledge, and solving simple math word problems.

- FACES finds substantial variation in the skill levels children have reached by the end of their time in Head Start. These variations have more to do with the socioeconomic and ethnic composition of the population served than with differences in teacher qualifications or program quality.

- FACES finds considerably less variation in the cognitive gains children make from fall to spring of Head Start than in the skill levels they achieve. These gains are less well-explained by either demographic characteristics of the population served or program quality indicators.

- FACES has found only weak or non-existent relationships between traditional measures of early childhood program quality and the gains Head Start children make in either the cognitive or social-emotional realms. The possible reasons for the lack of a relationship include: the general good quality of Head Start classrooms; the limited range of variation in quality constrains the magnitude of the possible relationships; and classroom quality may be a necessary but not a sufficient condition for practically significant gains in specific cognitive or behavioral areas.

The major implications of FACES suggest that increasing teacher qualifications and improving general classroom quality will not necessarily result in greater school readiness gains for Head Start children. What seems to be needed is the system-wide adoption of instructional methods with proven effectiveness in boosting specific areas of school readiness, such as "outside-in" skills.

A data set from the first two cohorts of FACES is now available for secondary analysis so these data can be explored further.

***Questions and Comments***

- *What norms are used?*
  The norms used in FACES are the general population norms developed by the test publishers. The issues of whether the norms are representative and whether they have become dated need to be addressed given the demographic changes taking place in the country and increasing racial/ethnic populations.

- *What is the impact of teacher education level on child learning?*

The FACES findings relating to teacher education level have been variable. The strongest implications from FACES is that the first emphasis should be on the implementation of instructional methods with proven efficacy in boosting some of the important readiness skills, and much of this can be done by teachers who do not have a B.A.

- *Were there differences across teachers based on education leve*l?
  FACES did not find significant differences in child learning outcomes relating to whether teachers had an A.A. or a B.A. degree. FACES does explore teacher qualities to a limited extent, but more research in this area is needed.

- *Were any relationships between the curriculum used and outcomes identified?*
  FACES looked at the two most popular curricula, High/Scope and Creative. In FACES 2000 there was a modest relationship between High/Scope and children making greater gains in the letter knowledge area, but this has not been found in FACES 2003. It is clear that programs using these two integrated curricula seem to be higher quality and have other characteristics that are desirable in Head Start programs. FACES does not specifically look at promising interventions so some of the most promising curricula that are being used are underrepresented in the sample.

- *Is a measure of parental literacy included in the study?*
  A direct measure of parental literacy is included in the study. There is a relationship between parental literacy and home activities, and the gains children made.

- *Are children with disabilities assessed as part of FACES and the NRS?*
  It is mandated that disabled children make up10 percent of the Head Start program. Both FACES and the NRS assess children with disabilities. The gains made by these children are comparable to those made by non-disabled children, but they still remain behind. The NRS indicates the children with disabilities are about six months behind non-disabled, English-speaking children from white families.

- *To improve outcomes, should more emphasis be place on instructional approach?*
  Teacher training and educational background may not be the most critical pieces in improving outcomes. Instructional methods may be more important. NRS collects information about the curriculum used by programs, but it does not assess how well the curriculum is being implemented.

## Overview of the Quality Research Centers (QRC) Consortium

*Dr. Ruth Hubbell-McKey, Xtria, LLC*

The mission of the QRC Consortium is to support the continuous improvement of Head Start by developing, testing, refining, and disseminating interventions to enhance the school readiness of Head Start children. The project officers are Lisa Trivits, Ph.D., and Maria Woolverton, Ph.D. The Data Coordinating Center (DCC) is headed by Nicholas Zill, Ph.D., at Westat. QRC studies are being conducted by the following grantees: Columbia University; Education Development Center; High/Scope Educational Research Foundation; Quality Counts, Inc.; State University of New York at Stony Brook; Temple University; University of North Carolina at Chapel Hill; and University of Oregon. A core set of data is collected from all the sites by the DCC.

The design for the consortium is as follows:

- Year 1 – Pilot year to test interventions in Head Start programs with a pre/post design;
- Year 2 – Test the interventions using common measures with a treatment/control design;
- Years 3–5 – Replicate interventions with new Head Start partners with a treatment/control design and common measures, and analyze and disseminate data.

The QRC data instruments follow the FACES model. The assessment battery includes: PPVT-III; McCarthy Draw-a-Design; letter naming task; Leiter-R Attention Sustained; color names and counting; WJ-R Letter-Word ID; WJ-R Applied Problems; WJ-R Dictation; and story and print concepts.

Each teacher rates each child in the sample. The items in the teacher-child report include: preschool learning behavior scale; child's disability status and identification process; COR rating items (problem solving, social relationships, creativity, music/movement, language/mathematics); behavior problems scale (aggression, withdrawal, hyperactivity); and classroom cooperative behavior scale.

Parent interviews are also included. The interviews must include core items and may be expanded to include additional optional items. The interviews address: activities with child, disability status, child's accomplishments, child's behavior, household rules and parenting practices, family configuration and demographics, child care, health and safety practices, community services, kindergarten transition, social support, depression and mastery, and parent involvement and satisfaction with Head Start.

Teacher interviews cover the following topics: education, experience, credentials, and membership; teacher beliefs scale; curriculum and classroom activities; training and mentoring; child assessment plans; job satisfaction; and management questions.

Classrooms are also observed by Westat. The observations include: counts of children and adults for determining ratio; Assessment Profile Scheduling; Assessment Profile Learning Environment; Assessment Profile Individualizing (2000); Teacher-Directed Activities Checklist; ECERS-R; and Arnett Caregiver Interaction Scale.

*QRC Programs*

*Using Assessment to Improve School Readiness and Head Start Program Quality*
National Center for Children and Families, Teachers College, Columbia University

Objective: To improve the school readiness of Head Start children, the quality of Head Start programs, and families' involvement in the program through the use of a nested observational assessment system.

Intervention: Train Head Start staff to administer, interpret, and use the results of the following assessments: child assessment (Work Sampling for Head Start); classroom assessment (Snapshot, ECERS-R, CLASS); and center assessment (Early Childhood Work Environment Scale). Intervention coordinators provide ongoing professional development and technical assistance at all three levels.

*System Approach to Fostering Language and Literacy Development*
Education Development Center, Inc.

Objective: To enhance classroom literacy outcomes, enhance child language and literacy outcomes, and build program capacity to sustain the training and quality improvement of the Head Start program.

Intervention: Program-Delivered Literacy through Inservice Training (PD-LIT) modules include analyzing and supporting children's writing, extending conversations, phonological awareness, print in the classroom, book reading, and integrating language and literacy into the curriculum.

*Achieving Head Start Effectiveness through Intensive Curriculum Training*
High/Scope Educational Research Foundation

Objective: To provide intensive training in, and confirm the practice of, the High/Scope curriculum to enable Head Start teachers to enhance children's development, especially in the areas of language, literacy, and the ability to resolve social conflict.

Intervention: Provide Head Start teaching staff with the High/Scope Preschool Curriculum Course training (20 days), and feedback and discussion sessions (five days).

*Preschool Behavior Project: a Socio-Emotional Intervention to Enhance School Readiness*
University of North Carolina at Chapel Hill

Objective: To reduce behavior problems and improve socio-emotional functioning by enhancing overall classroom quality, specific classroom strategies, existing mental health services, and parent involvement.

Intervention: The Preschool Behavior Project includes training and ongoing support for Head Start staff with pre-existing and newly developed curricula, including the Teachers and Children Series, Second Step, and Dialogic Reading; materials about relationship building and the behavioral change process; and group meetings for parents. The project includes services for universal and targeted groups.

*Head Start Adaptation of First Step to Success: Preparing Children for Social-Emotional Success in School*
Institute on Violence and Destructive Behavior, College of Education, University of Oregon

Objective: To improve school readiness, reduce serious behavior problems, and improve children's relationships with parents, teachers, and peers, through adapting the First Step to Success for use in Head Start.

Intervention: First Step to Success is designed to target specific children showing high rates of antisocial behavior. It is a collaborative home and school intervention of 45 to 60 days and is delivered by a behavior coach.

*Individualized Learning Intervention*
Quality Counts, Inc.

Objective: To enhance Head Start program quality through the use of assessments, a mentor program, and collaborative support to promote children's school readiness.

Intervention: The intervention has three components: 1) use of assessment information (classroom observations and developmental assessment); 2) self-directed learning experiences (mentor and protégé training); and 3) collaborative support of others (systemic change and a QRC Advisory Board).

*The Enhancement of Emergent Literacy Skills in Head Start: Outcomes of Classroom Curriculum Research*
State University of New York at Stony Brook

Objective: To identify curriculum approaches that provide effective enhancement of emergent literacy for children in Head Start.

Intervention: Two curriculum approaches, *Let's Begin with the Letter People* and *Waterford's Early Reading Program – Level 1*, will be studied. The High/Scope method will be used for comparison.

*The Companion Curriculum: Connecting Head Start Parents and Teachers to Promote Early Learning and Development*
Temple University

Objective: To improve parental involvement, parental satisfaction, and children's school readiness outcomes through implementing The Companion Curriculum (TCC).

Intervention: TCC consists of home-based educational activities promoting parent-child interaction that extend learning beyond the classroom; monthly educational workshops conducted by teachers; "Family Corner" established in classrooms to provide parents with a place to engage in readiness activities with children; and monthly teacher training regarding family involvement.

More information on the QRC is available in the *QRC 2004 Interim Report on Cross-site Analyses* and on the QRC Web site at:
< http://www.acf.hhs.gov/programs/opre/hs/qrc_two >.

## Head Start Impact Study (HSIS)

*Ronna Cook, Westat*
*Mike Puma, Westat*

As part of the Head Start reauthorization in 1998, Congress determined that HHS should conduct a national study to determine the impact of Head Start on the children served by the program. The impetus was concern about the lack of rigorous evidence regarding the impact of Head Start on participating children and families. Congress directed that the study be nationally representative and that it compare Head Start children with a group of comparable non-participants. The study team for the HSIS includes Westat, Chesapeake Research Associates, the Urban Institute, the American Institute for Research, and Decision Information Resources, Inc. An interim report containing data from the first year of the study was released in June 2005.

The study explores the impact of Head Start on key outcomes in four domains: cognitive; social-emotional; parenting practices; and children's experiences, which addresses the quality of educational settings and services. Health status is also included in the study. The research questions address both direct and indirect effects.

For direct effects, the research questions are:

- What difference does Head Start make to key outcomes of development and learning (and in particular, the multiple domains of school readiness) for low-income children?
- What works for which children and families?

For indirect effects, the research questions are:

- What difference does Head Start make to parental practices that contribute to children's school readiness?
- Does Head Start affect the nature of children's early care experience? Under what circumstances does Head Start achieve the greatest impact and what Head Start services are most related to impact?

To obtain comparable study groups, newly entering Head Start children were randomly assigned to either a treatment group or a control group. Children in the treatment group were enrolled in Head Start; children in the control group were not. For the control group, parents either found other services for their child, or the child was cared for at home. For ethical reasons, randomization was conducted only in sites where there were more applicants than slots available. Building partnerships at the local level was very important to the implementation of the study since there were serious questions about whether the study was feasible and whether programs would be willing to participate.

To ensure that a nationally representative sample was obtained, geographic clusters were created, and a stratified random sample of grantees was selected from the 25 clusters.

From the grantees in the stratified random sample, 383 centers were recruited for the study. Within the centers there was still variation—the most significant variation being whether the children were in a half-day or a full-day program.

All the students were randomly assigned, with 2,783 children in the treatment group and 1,884 in the control group. Due to saturation, approximately 15 percent of the children served by Head Start nationally are not included in the study.

In addition to assessing the two study groups, the study also explored the impact of Head Start on subgroups within the study population. For children, additional analysis looked at gender, race/ethnicity, language, incidence of special needs, and baseline score on outcome measures. Depressive symptoms, marital status, and whether the mother was a teen at first birth were the subgroups for parents.

Data sources used in the study are similar to those used in FACES. These include: child assessment; parent/primary care giver interview; teacher survey; care provider interview; center director interview; teacher/care provider child report; care setting observation (preschool years); and secondary information of schools.

Various data collection methods were used and data collection is overseen by local site coordinators. A major focus was determining which language to use to assess children. Parents report language-related information on the Head Start application is used. In addition, a Language Decision Form was developed that was administered by teachers. The form includes the following three questions:

- What is the language that the child speaks at home?
- What is the language that the child most often uses in the setting?
- What language do you think the child prefers to speak?

There was a very high correlation between the results of the Language Decision Form and the parent report.

Response rates have been good. The overall response rate was 80 percent in fall 2002, 83 percent in spring 2003, 84 percent in fall 2003, and 81 percent in spring 2004. The differentiation between the groups is decreasing over time. Locating the families in the control group over the course of the study has presented some challenges, especially now that some of the children are in kindergarten. Another significant challenge was assessing the children in the diverse settings. Children were assessed in centers, family day care, and in the home.

Spring 2005 data collection is currently being completed. In fall 2005, parent tracking will be conducted and in spring 2006, the 3 year-old cohort will be completed. The final report is expected in early 2007.

*Study Findings*

Preliminary estimates of the impact of Head Start, based on the first year of the study, are available, and analysis of the data continues. It is important to remember that when no impact is shown, this is not evidence that the program is not working. What it means is that the researchers found no evidence of whether there is an effect or not. This is not the same as finding clear evidence of no effect.

Year One impacts on children are listed below.

- Cognitive Domain – For both the 3 and 4 year-old cohorts, there were small to moderate positive impacts on pre-reading, pre-writing, and parent reports of children's literacy skills. There was no significant impact for oral comprehension, phonological awareness, and early mathematics skills.

- Social-Emotional Domain – For the 3 year-old cohort, there were small statistically significant impacts on reducing reported problem behaviors. There were no statistically significant impacts on social skills, approaches to learning, or on social competencies. For the 4 year-old cohort, there were no significant impacts in this domain.

- Parenting Practice Domain – For the 3 year-old cohort, there were small positive impacts on parents reading to their child and involvement in enrichment activities, a small impact on reduced use of physical discipline, and no significant impacts for safety practices. For the 4 year-old cohort, there were small impacts on parents reading to their children and no impacts for physical discipline or safety practices.

- Children's Experiences Domain – For both age cohorts, Head Start children were about twice as likely to use a center-based program. Non-Head Start children were about five times more likely to be exclusively in parent care. Head Start children were more likely to be in the same setting in both fall 2002 and spring 2003.

- Health Domain – For the 3 year-old cohort, there were small to moderate impacts on parent reports of children's access to health care and children's health status. For the 4 year-old cohort, there were moderate impacts on access to health care, but no significant impacts for health status. For both cohorts, there was also a large impact in access to dental care.

In comparisons between Head Start and non-Head Start centers, the findings indicate that children in Head Start centers were in environments that had more positive ratings of child-teacher interaction, more often used an instructional curriculum and activities to enhance children's skills, and had higher scores on the Early Childhood Environment Rating Scale, Revised Edition (ECERS-R).

*Planned Analysis*

Impacts on Children

- Estimate impacts for end of kindergarten and first grade for the 4 year-old cohort and estimate impacts for the end of the second preschool year, kindergarten, and first grade for the 3 year-old cohort. Where appropriate, this will include an analysis of impact on growth trajectories.
- Expand the set of child outcome measures to reflect the children's age.
- Add teacher-reported measures of social-emotional development and approaches to learning.
- Develop composite indices of child development and estimate impact on these higher-level aggregate measures.
- Test for difference in impact between the two age cohorts.
- Examine the impact of one vs. two years of Head Start.

Impacts on Parents

- Estimate impacts for end of kindergarten and first grade for the 4 year-old cohort, and at the end of second preschool year, kindergarten, and first grade for the 3 year-old cohort.
- Expand the set of parent outcome measures to include summer learning activities, school communication and involvement, and access to community services.
- Test for the difference in impact between the two age cohorts.
- Examine the impact of one vs. two years of Head Start.

Impacts on Children's Experiences

- Richer description of early care and school experiences of both age cohorts of children including: teacher qualifications; quality as defined by the ECERS-R score; lead teacher behavior as measured by Arnett; half-day vs. full-day programming; receipt of comprehensive services; activities in the classroom; and exposure to different types of child care experiences.
- Track year-to-year impacts on children's experiences.
- Test for the differences in impact between the two age cohorts and examine the impact of one vs. two years of Head Start.

The researchers are also developing a strategy for a non-experimental analysis of the relationship between impacts and the children's experiences. Three strategies are under consideration:

> Strategy 1 – Qualitative comparison of the experimental impacts on children's experiences, and impacts on child and parent outcomes.
>
> Strategy 2 – Examine the correlation between impact and site-by-site differences in experiences.
>
> Strategy 3 – Use "propensity scores" to create a synthetic comparison group.

***Questions and Comments***

- *Do findings from the first report indicate that Head Start Programs are having an impact?*
  Some positive effects have been found, and some promising effects have been found. Since this is only the first year of the study, it is not known if the effects will hold over time. These findings can help programs focus on what is being done well and then build on that to improve other areas.

- *Does the HSIS explore the cost per child of service?*
  The study does not have a cost component, and the collection of this kind of information is very difficult. However, information on cost and resources is critical when comparing programs and is necessary to thoroughly assess the impact of Head Start.

- *What is a reasonable effect size to expect?*
  Members discussed a reasonable effect size and what effect sizes should be expected. It was noted that significant policy decisions have been made in response to fairly small effect sizes. For example, studies of class size in which the effect size was fairly small have resulted in significant policy change relating to reduction of class sizes.

  Effect sizes depend a lot on the instruments used. Unless the assessment is totally adaptive, there will be some floor and ceiling effects, which will affect the effect size.

- *Are the children in the control group untreated?*
  With the study, some of the children in the control group are not untreated. Children in both groups are receiving treatment so it could be expected that effect sizes might be small. However, it will be important to look carefully at the settings for non-Head Start children. Many are in homes or subsidized daycare. This should not be considered treatment.

## PRISM and Local Assessments

*Dr. Michele Plutro, Head Start Bureau*

PRISM is both an instrument (with a set of protocols) and the process used to conduct Federal monitoring of Head Start and Early Head Start programs. Monitoring of Head Start programs is important in that it helps strengthen services provided to and for enrolled children and their families. It ensures that they are receiving quality services and provides grantees with information, in the form of detailed reports, which can be used in program improvement efforts.

While the monitoring can assist with program improvements, it also must be done to comply with the statutory mandate that all Head Start grantees be monitored once every three years. If deficiencies are identified during the review process, additional levels of review may be required. Newly designated agencies are reviewed at the completion of their first year. The Secretary also has the discretion to conduct onsite visits for any purpose. This usually occurs when ACF is notified about serious issues in a program. Termination of grants is rare, but the number of grants terminated is increasing.

PRISM employs a systems approach to monitoring, focusing on how a grantee's systems and services interact to create and sustain a quality program. Prior to the development of the systems approach, the monitoring looked primarily at service delivery.

The reviews are conducted by teams. Each team is supervised by a team leader. There are both systems and service reviewers. Systems reviewers examine information related to management systems and program governance. System reviewer expertise includes: family and community partnerships, program design and management, report writing and coordination, fiscal management, and facilities. Service reviewers examine grantee service implementation and partnership activities. Areas of service reviewer expertise include: early childhood development, health, mental health, disabilities, infant and toddler development, infant and toddler disabilities, and infant and toddler maternal and child health. All reviewers examine inter-relationships between services and systems.

The PRISM process is designed to monitor grantee performance against program requirements. Head Start standards are at the core of this process. The reviewers examine the extent to which the grantee is in compliance with program requirements. Reviewers collaborate to examine whether identified problems reflect larger management or other problems within the system as a whole.

The PRISM instrument is designed to organize standards into a manageable framework and serves as a tool to facilitate and organize the data collection process. Tools within the PRISM instrument include: 18 core questions; guidance on how to conduct reviews; and review protocols, such as checklists and instruments for recording observations. The core questions, half of which apply to systems and half to services, address the following: program governance; planning; communication; record keeping and reporting; ongoing monitoring; program self assessment; human resources; fiscal management; prevention

and early intervention; individualization; mental health; disabilities services; curriculum and assessment; family partnership building; parent involvement; community partnerships; eligibility, recruitment, selection, enrollment and attendance; and facilities, materials, equipment, and transportation. For each of the core questions, there is a very specific body of requirements that each Head Start program must address. Programs have flexibility in how they will meet these requirements; ACF does not dictate how it must be done. A key element of PRISM is the emphasis on the interrelated nature of the core questions, especially across the service and systems areas.

There is a great deal of variation across grantees in terms of their assessment ability. Some are very sophisticated and others cannot differentiate between screening and ongoing assessment, or how to use screening and assessment as program planning tools. Grantees are required to include regular assessments of children as part of their overall assessment process. In addition, programs must also solicit observations and comments from parents. Many grantees are now using commercial tools in their assessment process. The two most popular commercial tools are associated with the High/Scope and Creative curricula. Both these tools are capable of producing computer-generated reports at the child, classroom, center, or program level. The Advisory Committee may find it useful to hear from some grantees about their ongoing assessment process.

Additionally, the core questions are framed to prevent programs or reviewers from becoming narrowly focused on a small aspect or key word within the core question so that they miss the big picture. Within all the core questions, there are nuances or tangential relationships established between the various core questions. For example, in core question five, which relates to ongoing monitoring—programs are asked what they know about the status of the children in the program and then aggregate what they know. This requirement to aggregate data has been greatly responsible for the increase in commercial assessment tools. The aggregation at the program level is carried out in various ways. Some programs aggregate the frontline evidence on a set of tools. Other programs aggregate only at the classroom level. There can be parallel systems within an agency.

Data from the review process are entered into the PRISM database, which is used to generate review reports and annual reports to Congress. The final product of the PRISM process is a review report, which is finalized by ACF and sent to the grantee. The review team provides evidence to ACF. The interpretation of the evidence is done by ACF.

Unfortunately, the PRISM database is not integrated with other ACF databases, which include PIR, HSMTS, FACES, and QRC. There is some duplication across these databases but they also include unique data. Efforts to create a platform have failed so it will be at least another year before the various databases are linked.

### *Questions and Comments*

Answers to questions related to curricula for Head Start programs.

- High/Scope and Creative were developed in parallel with the Head Start program. The performance standards that were revised in 1998 require that grantees have a written curriculum. Curricula have to have clear goals for children's development and learning, describe specific types of experiences for children, identify clear roles for teachers and parents, identify specific materials and equipment, conform to principles of sound child development, and contain performance standards.

- Any standard curriculum will need to be adapted at the local level. In addition, how the curriculum is implemented at the local level will have more impact on outcomes than the curriculum itself. If a great curriculum is poorly implemented, the outcomes will probably be poor.

- It appears that PRISM and the performance standards may be driving curriculum selection.

- There are few studies available that demonstrate the efficacy of specific curricula but there are currently studies that are comparing curricula.

## Recommendations: Areas and Issues for Consideration

1.) Coordination of Assessment Activities

- There are multiple Head Start-related assessment activities. There is a lot of overlap, but the activities are not necessarily redundant.
- Ensure that assessment activities are complementary.
- Minimize duplication across assessments.
- Determine if more guidelines relating to assessment are necessary for programs.

2.) Expansion and Enhancement of NRS

- Explore use of NRS at the center, classroom, and individual level so that data can be used to inform and enhance program performance.
- Expand NRS to include social development, chronological awareness, and more parental assessment.
- The NRS was not designed to provide class and individual data. In designing the assessment, it was kept short intentionally to minimize the burden to programs and students.

3.) Curriculum

- Should there be greater use of curricula that have been proven effective?
- Is a national curriculum the answer?

4.) Parental Involvement

- Explore interventions to engage parents and support them in helping their children learn.
- Review existing data relating to parent involvement and satisfaction in FACES and other data sets.

5.) English-Language Learners

- Disaggregate data on English-language learners from existing studies (FACES, NRS).
- English language learners show large gains, but they are not large enough to bring them up to the level of children from English speaking families.
- English-language learners start out so behind the other students that it is unrealistic to expect them to make huge advances in a relatively short period of time. Acceptable rates of growth should be determined.

- NRS data could be used to compare children with low skill levels in both Spanish and English with children who are English language learners with good Spanish skills.
- There is evidence that Spanish language acquisition predicts English attainment.
- Logistical concerns about administering assessments to this population exist, such as the availability of assessors who speak Spanish.

6.) Effect Size

- What is an acceptable effect size?
- How should effect size be determined?
- Are achievement goals also necessary?

7.) Teacher Qualifications

- Does the education level of teachers make a difference?
- How does teacher performance relate to outcomes?
- What teacher characteristics result in improved outcomes?

8.) NRS Administration Levels

- Who should administer the NRS?
- Is there variability based on who is giving the assessment?
- Does there need to be more standardization in the administration of the assessment?

9.) Children with Disabilities and the NRS

- What accommodations are made for children with disabilities in administration of the NRS?
- Do the accommodations affect performance on the assessment?
- Are current accommodations appropriate?
- Are all children with disabilities being assessed? Should they all be assessed?

10.) Assessment of Early Head Start Programs

- Should the NRS be extended to Early Head Start?

## Additional Information/Resources Requested for Future Meetings

1.) TWG Meeting Outcomes

- Slides and other materials;
- Meeting summary;
- *These materials are available from Westat.*

2.) HSIS, FACES, and NRS Commonalities

- Develop a chart or interface comparing HSIS, FACES, and NRS.
- Include key elements and study findings (what the studies reveal).
- *To be developed by Westat.*

3.) Response of Programs to NRS

- Explore how programs perceive the NRS (assessment process, burden to program, data availability, and usefulness of the data to programs).
- Provide information on the time and other resources that programs expend on NRS.
- Data on the programs' perceptions of and satisfaction with the NRS are available from Mathematica surveys and the record of public commentary.

4.) Outcomes from curriculum efficacy studies and other studies relevant to Head Start Programs

- National Early Literacy Panel (NELP)
- Preschool Curriculum Evaluation Research (PCER) Program
- What is the current state of evidence on interventions and their impact on student outcomes?
- Identify predictors of specific outcomes (e.g., late reading).

5.) Elements of Program Success

- Compare programs with large gains to programs with smaller gains.
- Data are available from the PIR and the CBRS. Characteristics that can be explored include demographics of children, disability, teacher education level, operational aspects (full vs. half-day), etc.

6.) National Institute of Child Health and Human Development (NICHD) School Readiness Research

7.) Early Childhood Longitudinal Study (ECLS)

8.) Information on other ACF assessment activities including CORS and the university partnerships to develop new measures

9.) Social-Emotional Development Instrument

Committee members requested to see the draft instrument that Westat will use for the teacher-conducted assessments of social-emotional development that are scheduled to be conducted in fall 2005 and spring 2006.

## Ethics Review

Individuals appointed to serve on HHS advisory committees or Presidential boards, councils, or commissions are "special Government employees" (SGEs), which are defined as "an officer or employee in the executive branch of the Federal Government who is appointed to perform temporary duties, with or without compensation, for a period not to exceed 130 days during any period of 365 consecutive days," 18 U.S. C. § 202(a). HHS advisory committee members appointed as SGEs are required under the Ethics in Government Act, as amended by the Ethics Reform Act of 1989, and 5 C.F.R. Part 2634, to file a financial disclosure report when first appointed and annually thereafter on the anniversary date of their appointment. Committee members also may be required to update the information on the report before each meeting through their term of appointment. The information reported is used to determine the matters for which a committee member must be disqualified under the criminal financial conflict of interest statute, 18 U.S. C. § 202(a), and the matters for which a committee member may be granted a waiver under 18 U.S. C. § 202(b).

SGEs must complete certain forms. These include the Confidential Financial Disclosure Report (SF-450) to determine possible conflict of interest. Conflicts are resolved primarily through recusal waiver. SGEs are also required to complete the HHS Form 697, the Foreign Activities Questionnaire, to determine if foreign connections are appropriate. The U.S. Constitution Emoluments Clause prohibits an SGE from having an employment relationship with a foreign government. The Foreign Gifts and Decorations Act specifies that SGEs can only accept gifts of up to $305 from foreign governments.

Complete reporting is essential to protect the committee member from inadvertently violating any of the criminal conflict of interest statutes, and to assure the public that the advice provided by an HHS advisory committee is free from any real, or perceived, conflicts of interest. The information reported by committee members is confidential and may not be released except under limited circumstances.

Numerous statutes apply to SGEs but one of the most significant is ***18 U.S.C. § 208.*** Section 208(a), the main conflict of interest statute, prohibits an SGE from participating personally and substantially in any particular matter that could affect the financial interests of the SGE, the SGE's spouse, minor child, general partner, an organization in which the SGE serves as an officer, director, trustee, general partner, or employee, or an organization with which the SGE is negotiating or with which the SGE has an arrangement for prospective employment. Under this statute, for example, an SGE would be prohibited from reviewing a grant application submitted by a researcher from the same university in which the SGE is employed. The SGE would be recused from participation in the review.

Section 208(b)(3) authorizes issuance of a waiver to an SGE who serves on a committee subject to the Federal Advisory Committee Act if the official responsible for the individual's appointment certifies in writing that the need for the individual's services

outweighs the potential for a conflict of interest created by a particular financial interest involved. The waiver granted is considered a "general" waiver, in that it allows participation in matters that affect all institutions, or types of institutions, similarly. Even with a general waiver, however, SGEs must disqualify themselves from participation in all matters that specifically and uniquely affect their financial interests.

In addition, under regulations issued by the Office of Government Ethics, a regulatory (i.e., automatic) waiver of the disqualification requirement of 18 U.S.C. § 208 is available under certain circumstances, including instances involving the following classes of financial interests:

- Interests held in broadly diversified investment funds;
- Publicly traded securities of $15,000 or less;
- Publicly traded securities of $25,000 or less if the matter is a general policy matter and the total value of all investments in the affected industry sector is no more than $50,000;
- Employment in one campus of a multi-campus State university if the matter affects only another campus and the employee does not have multi-campus responsibilities.

In addition, there is an automatic exemption which allows SGEs serving on Federal advisory committees to participate in particular matters of general applicability where the otherwise disqualifying financial interest arises solely from the committee member's non-Federal employment or prospective employment, provided that the matter will not have a special or distinct effect on the employee or employer other than as part of a class. This exemption is unavailable if the employee (or those persons whose interests are imputed to the employee) owns stock, stock options, or has some other financial interest in the employer other than his or her employment interest.

Other important criminal conflict of interest statutes are listed below.

***18 U.S.C § 201***. Section 201, commonly known as the "bribery and illegal gratuities" statute, prohibits Federal employees, including SGEs, from seeking, accepting, or agreeing to receive anything of value in return for being influenced in the performance of an official act.

***18 U.S.C. § 203***. Section 203 prohibits an SGE from receiving compensation for representational services rendered by the employee or another person before HHS or another Federal agency or other specified entity (such as a court or commission) in any particular matter involving a specific party in which: 1) the SGE has participated personally and substantially as a Government employee; or is 2) pending in the Government agency in which the SGE is serving.

***18 U.S.C. § 205.*** Section 205 prohibits an SGE from representing a party, with or without compensation, before HHS or another Federal agency or other specified entity (such as a court or commission) in any particular matter involving a specific party in which the

United States is a party or has a direct and substantial interest: 1) that the SGE participated in personally and substantially as a Government employee; or 2) which is pending in the agency in which the SGE is serving, if the SGE has served for more than 60 days during the immediately preceding 365 days.

***18 U.S.C. § 207.*** Section 207, the "post-employment" statute, imposes a lifetime ban on a former SGE from representing another person or entity to HHS or another Federal agency or other specified entity (such as a court or commission) in any particular matter involving a specific party in which the former SGE participated personally and substantially while serving in the Government. In addition, for two years after terminating Federal employment, an SGE may not make such representational communications to the Government regarding specific party matters that were pending under his or her official responsibility during the last year of Government service.

***18 U.S.C § 219***. Section 219 prohibits an SGE from acting as an "agent of a foreign principal" as defined under the Foreign Agents Registration Act (FARA) or a "lobbyist" on behalf of a foreign entity that is required to register under the Lobbying Disclosure Act of 1995 (LDA).

Additional regulations that apply to SGEs are listed below.

*Teaching, speaking, and writing in a personal capacity*
An SGE is prohibited from receiving compensation for teaching, speaking, and writing that relates to the employee's duties. An SGE may receive compensation for teaching, speaking, or writing in a personal capacity. For SGEs serving less than 60 days, the restriction only applies to particular matters involving specific parties.

*Gifts*
An SGE may not accept gifts offered as a result of his or her membership on an advisory committee.

*Charitable Fundraising*
An SGE may engage in charitable fundraising in a personal capacity as long as the SGE does not personally solicit funds or other support from any person or entity known to him or her to have interests that may be substantially affected by the performance or nonperformance of the SGE's Federal duties.

*Expert Witness*
An SGE may not participate as an expert witness in connection with any matter or proceeding that the SGE works on as an SGE.

*Impartiality*
An SGE is prohibited from participating in a specific party matter where a reasonable person would question the SGE's impartiality.

*Misuse of Position*

An SGE may not use his or her position to imply that the Advisory Committee or HHS endorses his or her activities and may not refer to his of her Government position for his or her own private gain.

*Lobbying*
An SGE is prohibited from engaging in any activity which directly or indirectly encourages or directs any person or organization to lobby one or more members of Congress. An SGE may appear before any individual or group for the purpose of informing or educating the public about a particular policy or legislative proposal.

*Political Activity*
The Hatch Act prescribes the restrictions on certain political activities of Federal employees. Hatch Act restrictions only apply during the period of any day in which the SGE is actually performing Federal Government business.

Ethics contacts for the Administration for Children and Families are listed below.

Curt Coy, Deputy Ethics Counselor, 202/402-9238
Donnell Savage, Ethics Contact, 202/401-4797

Ethics Division, Office of the General Counsel

Edgar M. Swindell, Designated Agency Ethics Official, 202/690-7258
Michael Wolf, Ethics Counsel, 202/260-1792